DOI:
10.1107/S2059798316005830

Document Version
Peer reviewed version

STRUCTURAL BIOLOGY

ISSN: 2059-7983
journals.iucr.org/d

A crystallographic study of human NONO (p54^{nrb}): overcoming pathological problems with purification, data collection and noncrystallographic symmetry

Gavin J. Knott, Santosh Panjikar, Andrea Thorn, Archa H. Fox, Maria R. Conte, Mihwa Lee and Charles S. Bond

Acta Cryst. (2016). D**72**, 761–769

STRUCTURAL BIOLOGY
ISSN 2059-7983

A crystallographic study of human NONO (p54^{nrb}): overcoming pathological problems with purification, data collection and noncrystallographic symmetry

Gavin J. Knott,[a] Santosh Panjikar,[b,c] Andrea Thorn,[d] Archa H. Fox,[e,f] Maria R. Conte,[g] Mihwa Lee[h] and Charles S. Bond[a]*

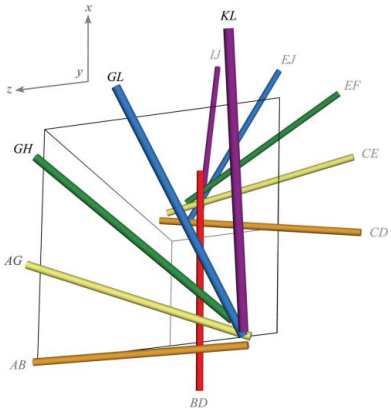
[a]School of Chemistry and Biochemistry, The University of Western Australia, Crawley, WA 6009, Australia, [b]Australian Synchrotron, Victoria, Australia, [c]Department of Biochemistry and Molecular Biology, Monash University, Melbourne, VIC 3800, Australia, [d]MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge Biomedical Campus, Cambridge CB2 0QH, England, [e]School of Anatomy, Physiology and Human Biology, The University of Western Australia, Crawley, WA 6009, Australia, [f]Harry Perkins Institute of Medical Research, QEII Medical Centre, Nedlands, WA 6009, Australia, [g]Randall Division of Cell and Molecular Biophysics, King's College London, New Hunt's House, Guy's Campus, London SE1 1UL, England, and [h]Department of Biochemistry and Genetics, La Trobe Institute for Molecular Science, La Trobe University, Melbourne, Victoria, Australia. *Correspondence e-mail: charles.bond@uwa.edu.au

Received 2 February 2016
Accepted 8 April 2016

Edited by T. O. Yeates, University of California, USA

Keywords: NONO (p54^{nrb}); L-proline; cumulative intensity distribution; *L* statistic; noncrystallographic symmetry.

Non-POU domain-containing octamer-binding protein (NONO, a.k.a. p54^{nrb}) is a central player in nuclear gene regulation with rapidly emerging medical significance. NONO is a member of the highly conserved *Drosophila* behaviour/human splicing (DBHS) protein family, a dynamic family of obligatory dimeric nuclear regulatory mediators. However, work with the NONO homodimer has been limited by rapid irreversible sample aggregation. Here, it is reported that L-proline stabilizes purified NONO homodimers, enabling good-quality solution small-angle X-ray structure determination and crystallization. NONO crystallized in the apparent space group $P2_1$ with a unique axis (*b*) of 408.9 Å and with evidence of twinning, as indicated by the cumulative intensity distribution *L* statistic, suggesting the possibility of space group *P*1. Structure solution by molecular replacement shows a superhelical arrangement of six NONO homodimers (or 12 in *P*1) oriented parallel to the long axis, resulting in extensive noncrystallographic symmetry. Further analysis revealed that the crystal was not twinned, but the collected data suffered from highly overlapping reflections that obscured the *L*-test. Optimized data collection on a new crystal using higher energy X-rays, a smaller beam width and an increased sample-to-detector distance produced non-overlapping reflections to 2.6 Å resolution. The steps taken to analyse and overcome this series of practical difficulties and to produce a biologically informative structure are discussed.



1. Introduction

The *Drosophila* behaviour/human splicing (DBHS) protein family are predominantly nuclear-localized proteins that coordinate protein and nucleic acid interactions throughout nuclear gene regulation (Knott *et al.*, 2016). An ensemble of domains commonly referred to as the 'DBHS region' defines DBHS proteins, comprising tandem N-terminal RNA-recognition motifs (RRMs), a NONA/paraspeckle (NOPS) domain and a C-terminal coiled coil (Dong *et al.*, 1993). In higher order mammals, the DBHS protein family constitutes the non-POU domain-containing octamer-binding protein (NONO or p54^{nrb}), splicing factor proline/glutamine-rich (SFPQ or PSF) and paraspeckle protein component 1 (PSPC1). DBHS proteins are well established as obligatory

dimers, and X-ray crystallography has illustrated the intimacy of dimerization with the structures of the PSPC1–NONO heterodimer, SFPQ homodimer and NONO-1 homodimer (Passon *et al.*, 2011, 2012; Lee *et al.*, 2015; Knott *et al.*, 2015). NONO can be found as a heterodimer with SFPQ or PSPC1 (Straub *et al.*, 1998; Fox *et al.*, 2005; Passon *et al.*, 2012) and has been reported to function in transcriptional co-activation and co-repression (Ishitani *et al.*, 2003; Dong *et al.*, 2005; Yadav *et al.*, 2014), post-transcriptional processing (Peng *et al.*, 2002; Kameoka *et al.*, 2004; Liang & Lutz, 2006; Hall-Pogar *et al.*, 2007; Kaneko *et al.*, 2007; Izumi *et al.*, 2014; Yadav *et al.*, 2014; Lu & Sewer, 2015), paraspeckle assembly and function (Fox *et al.*, 2002, 2005; Chen & Carmichael, 2009), DNA repair (Udayakumar & Dynan, 2015) and circadian rhythm (Kowalska *et al.*, 2012, 2013). However, emerging evidence indicates that NONO itself is a key regulator of the cell-cycle G1-S checkpoint (Kowalska *et al.*, 2013), viral infection (Cao *et al.*, 2015) and cancer (Zhu *et al.*, 2016). With a developing clinical significance (Cao *et al.*, 2015; St Gelais *et al.*, 2015; Mircsof *et al.*, 2015; Zhu *et al.*, 2016) and a broad diversity of annotated functions, methods for the effective expression and purification of NONO are desirable. Towards this goal, we report a protocol to successfully purify folded and stable NONO homodimer. Unlike other members of the DBHS protein family, we report that handling of NONO requires methods that limit the rapid aggregation of the purified protein. Furthermore, we describe the challenging X-ray crystallographic analysis of NONO crystals that provides insight into the effect of overlapping reflections on the cumulative intensity distribution L statistic (Padilla & Yeates, 2003). Despite the challenges, we determined the structure of the NONO homodimer by molecular replacement and detail its molecular packing and extensive noncrystallographic symmetry. The challenges encountered and discussed in this study highlight the importance of scrutinizing reflections and optimizing data-collection strategies.

2. Materials and methods

2.1. Expression and purification of the NONO homodimer

The DBHS region (residues 53–312 with respect to UniProt entry NONO_HUMAN) was cloned into pET-Duet-1 (Novagen) with an inserted TEV protease site (bold) (5′-CAGGATCC**AGAAAACCTGTATTTTCAGGGC**CATAGTGCATCCCTTAC-3′ and 5′-CGGAATTCTTACATTAGCATGACCTGGTG-3′; BamHI and EcoRI sites are underlined in the sequences). Competent *Escherichia coli* Rosetta2 (DE3) cells (Novagen) were transformed and selected from lysogeny broth (LB) agar plates (100 µg ml^{-1} ampicillin, 50 µg ml^{-1} chloramphenicol). Single colonies were inoculated into 5 ml LB [100 µg ml^{-1} ampicillin, 50 µg ml^{-1} chloramphenicol, 0.1%(*v*/*v*) D-glucose] and incubated for 16 h at 310 K and 200 rev min^{-1}. The 5 ml culture was used to inoculate 500 ml LB (50 µg ml^{-1} ampicillin, 50 µg ml^{-1} chloramphenicol, 0.1% D-glucose) in 2 l conical flasks incubated at 310 K and 200 rev min^{-1}. At an optical density (600 nm) of 0.6, the cultures were cooled on ice before induction with 0.5 m*M* IPTG for 16 h at 298 K and 180 rev min^{-1}. Compact pellets of 500 ml were gently resuspended on ice in 50 ml buffer [50 m*M* Tris–HCl pH 7.5, 250 m*M* NaCl, 25 m*M* imidazole, 10%(*v*/*v*) glycerol] supplemented with 1000 units of Benzonase nuclease (Sigma). Lysis was carried out with an Emulsiflex C5 high-pressure homogeniser (Avestin) and the lysate was clarified by centrifugation (24 000*g* for 30 min at 278 K) and 0.22 µm filtration before application onto a 5 ml $NiCl_2$-charged HiTrap column (GE Healthcare). NONO was eluted using a ten-column-volume imidazole gradient (25–500 m*M*). After an immediate 1:3 dilution in gel-filtration buffer [20 m*M* Tris–HCl pH 7.5, 50 m*M* L-proline, 250 m*M* KCl, 0.5 m*M* EDTA], NONO was incubated at ambient temperature for 16 h with recombinant His-TEV protease supplemented with 1 m*M* dithiothreitol. Post-digestion centrifugation (24 000*g* for 30 min at 277 K) and 0.22 µm filtration were carried out before re-application onto a HiTrap column (GE Healthcare). Cleaved NONO was pooled and loaded in 5 ml injections onto a HiLoad 16/60 Superdex 200 column (GE Healthcare) developed with gel-filtration buffer at 1 ml min^{-1}. NONO was concentrated after gel filtration using a 10 kDa concentrator (Amicon) and the concentration was determined from the absorption at 280 nm using an estimated absorption coefficient of 11 460 M^{-1} cm^{-1}. Purified NONO homodimer was stored at 277 K or flash-cooled with liquid nitrogen for long-term storage at 193 K. Gel-filtration experiments on NONO samples with and without L-proline were carried out using an S200 16/60 column (GE Healthcare) developed with gel-filtration buffer with or without L-proline at room temperature at 1.0 ml min^{-1}.

Table 1
SAXS data-collection and analysis statistics.

	NONO without L-proline	NONO with L-proline
Data collection		
Instrument	Australian Synchrotron SAXS/WAXS	Australian Synchrotron SAXS/WAXS
Strategy	Concentration series (static)	SEC-SY-SAXS
SEC column	N/A	WTC-030N5 (Wyatt)
Beam geometry (µm)	22	22
Wavelength (Å)	1.12713	1.03320
q range (Å^{-1})	0.010–0.490	0.006–0.375
Exposure per frame (s)	1.00	5.00
Flow rate (ml min^{-1})	0.5	0.2
Concentration (mg ml^{-1})	3.5	7.0 (at injection)
Temperature (K)	298	298
Guinier parameters		
$I(0)$ (cm^{-1})	0.03 ± 0.00	0.03 ± 0.00
R_g (Å)	44.63 ± 0.84	28.05 ± 0.67
Guinier range (q^2)	0.0001–0.0016	0.0001–0.0016

2.2. Small-angle X-ray scattering

Small-angle X-ray scattering data were collected on the SAXS/WAXS beamline at the Australian Synchrotron (Kirby *et al.*, 2013) with continuous data collection using a Pilatus 1M detector. For the NONO homodimer in the absence of L-proline, static measurements were collected in triplicate in

Table 2
NONO data-collection and processing statistics.

Values in parentheses are for the highest resolution bin.

	Data set 1a (13 keV)	Data set 1b (13 keV)	Data set 1c (13 keV)	Data set 2 (15 keV)
Space group	$P2_1$	$P1$	$P2_1$	$P2_1$
Unit-cell parameters (Å, °)	a = 67.31, b = 408.19, c = 69.21, β = 97.98	a = 67.31, b = 69.21, c = 408.19, α = 90.08, β = 89.81, γ = 97.98	a = 67.14, b = 408.6, c = 69.32, β = 98.03	a = 67.15, b = 407.18, c = 68.96, β = 97.75
Molecules per asymmetric unit	12	24	12	12
Resolution (Å)	48.3–2.95 (3.11–2.95)	48.3–2.95 (3.11–2.95)	48.3–2.95 (3.04–2.95)	48.1–2.60 (2.65–2.60)
Measured reflections	280114	280177	336813	412058
Unique reflections	76059 (10475)	146347 (20395)	55855 (4338)	110444 (5451)
Completeness (%)	98.8 (95.5)	95.9 (94.1)	72.4 (68.1)	98.9 (97.7)
Multiplicity	3.7 (3.7)	1.9 (1.9)	6.0 (6.1)	3.7 (3.8)
$\langle I/\sigma(I)\rangle$	8.3 (2.6)	7.6 (2.1)	5.0 (1.6)	10.1 (1.6)
$CC_{1/2}$	0.760	0.580	0.613	0.524
R_{merge} (%)	0.107 (0.455)	0.07 (0.360)	0.196 (1.02)	0.088 (0.742)
Average mosaicity (°)	0.24	0.24	0.25	0.11
Wilson B factor (Å^2)	66.01	69.21	100.66	36.58
L statistic	0.423	0.423	0.500	0.463
Potential twin operator	$-l$, $-k$, $-h$	$-h$, $-k$, l	—	—
Potential twin fraction	0.11	0.50	—	—

20 m*M* Tris–HCl pH 7.5, 250 m*M* KCl, 0.5 m*M* EDTA. For the NONO homodimer in the presence of L-proline, data were collected in gel-filtration buffer using size-exclusion chromatography-coupled synchrotron small-angle X-ray scattering (SEC-SY-SAXS; Gully *et al.*, 2015) controlled by a Shimadzu HPLC system. SAXS data-collection and analysis statistics are presented in Table 1. Scattering data in the absence of L-proline were background-corrected using linear interpolation of the background from averaged frames. Scattering data collected in the presence of L-proline were background-corrected using the *US-SOMO* SAS module (Brookes *et al.*, 2013). The Guinier range, the radius of gyration (R_g) and $I(0)$ were determined with the *ATSAS* software (Petoukhov *et al.*, 2012) using *PRIMUS* (Konarev *et al.*, 2003) and *GNOM* (Semenyuk & Svergun, 1991). The theoretical solution scattering of NONO was calculated and fitted to the experimental scattering using *CRYSOL* (Svergun *et al.*, 1995).

2.3. Crystallization and X-ray data collection

Crystallization space was explored using the Index screen (Hampton Research) set up in 96-well sitting-drop vapour-diffusion format by an Art Robbins Phoenix robot as follows: 150 nl protein solution (1.5–3.0 mg ml^{-1}) plus 150 nl reservoir solution was equilibrated against 80 µl reservoir solution in 96-well format ARI LVR Intelli-Plates (Hampton Research) at 293 ± 0.5 K. Clusters of thin plates formed in condition No. 44 of the Index screen, with 24-well sitting-drop optimization yielding larger crystal clusters using 4 µl protein solution (1.5 mg ml^{-1}) plus 4 µl reservoir solution [0.1 *M* HEPES pH 8.0, 20%(*w*/*v*) PEG 3350] equilibrated against 500 µl reservoir solution at 293 ± 0.5 K. Crystal clusters were fished out and transferred to cryoprotection solution [0.1 *M* HEPES pH 8.0, 20%(*w*/*v*) PEG 3350, 15%(*v*/*v*) glycerol], where Micro-Tools (Hampton Research) were used to separate the clusters before flash-cooling in liquid nitrogen. Diffraction experiments were carried out on the MX2 beamline at the Australian Synchrotron, Melbourne, Victoria, Australia. Data set 1 was collected at 13 keV (λ = 0.953 Å) with a crystal-to-detector distance of 350 mm (edge at 2.27 Å). An additional data set (data set 2) was collected at 15 keV (at λ = 0.826 Å) after introducing a 10 µm collimator and increasing the crystal-to-detector distance to 450 mm (edge at 2.46 Å). The differing data-collection strategies are illustrated schematically along with representative X-ray diffraction patterns for each data set from two orthogonal perspectives in Figs. 3(*a*) and 3(*b*).

2.4. Data processing and structure solution

Data-collection and processing statistics are presented in Table 2. Both data sets were processed and scaled in space group $P2_1$ using *XDS* (Kabsch, 2010) and *AIMLESS* (Evans & Murshudov, 2013), respectively. The cumulative intensity distribution of both data sets was analysed using the L statistic to indicate the presence of twinning (Padilla & Yeates, 2003). Detailed analysis of the processed data sets was also carried out using *phenix.xtriage* (Afonine *et al.*, 2012). Solvent-content analysis (Matthews, 1968) indicated the most likely asymmetric unit content to be between ten and 14 molecules for both data sets. The self-rotation function was calculated with *MOLREP* (Vagin & Teplyakov, 2010) using data between 40.0 and 5.0 Å resolution with an integration radius of 50.0 Å, revealing multiple significant peaks on the κ = 180° section. Space-group validation was carried out in *ZANUDA* (Lebedev & Isupov, 2014). Molecular-replacement calculations were carried out in $P2_1$ with *MOLREP* using the *Caenorhabditis elegans* NONO-1 dimer truncated to residues 115–345 (PDB entry 5ca5) as the search model (Knott *et al.*, 2015). Searches were conducted for five, six and seven dimers using default parameters to a maximum resolution of 5.0 Å. Initially, the highest correlation was obtained by placing four dimers of the truncated 5ca5 structure in the asymmetric unit. However, after rigid-body refinement with *REFMAC*5 (Murshudov *et al.*, 2011) the resulting electron-density map

indicated the presence of two additional dimers. The dimer array was completed using the fixed rigid-body solution and by searching for two additional dimers of the 5ca5 structure. The residues of the molecular-replacement model were substituted by the correct sequence for *Homo sapiens* NONO using *Coot* (Emsley *et al.*, 2010). Coordinates were manipulated with *PDB-MODE* (Bond, 2003). The structure was refined using *BUSTER* (Blanc *et al.*, 2004), *REFMAC*5 and *phenix.refine*. A partially refined model was used to evaluate the pathology of the data sets described here. The coordinates and structures factors of the fully refined model, validated using *MolProbity* (Chen *et al.*, 2010), have been deposited in the PDB (PDB entry 5ifm). The final refinement statistics along with accompanying structural and biochemical interpretations will be reported elsewhere. *LSQMAN* (Kleywegt, 1996) was used to calculate the rotation matrices and translation vectors that superimpose molecules in the asymmetric unit onto each other and the vectors describing the rotation axes between molecules.

3. Results and discussion

3.1. Purification and solution X-ray scattering

Constructs containing the DBHS region of NONO form obligatory homodimers that rapidly aggregate irreversibly over time and as a function of increasing concentration. Assessment by optimal solubility screening (Jancarik *et al.*, 2004) and differential scanning fluorimetry (DSF; Seabrook & Newman, 2013) using a broad range of buffers, pH, salts and additives did little to improve the behaviour of purified NONO. Addition of the cosmotropes L-arginine/L-glutamate (Golovanov *et al.*, 2004) or L-proline was then tested, with L-proline being found to remarkably improve the stability of

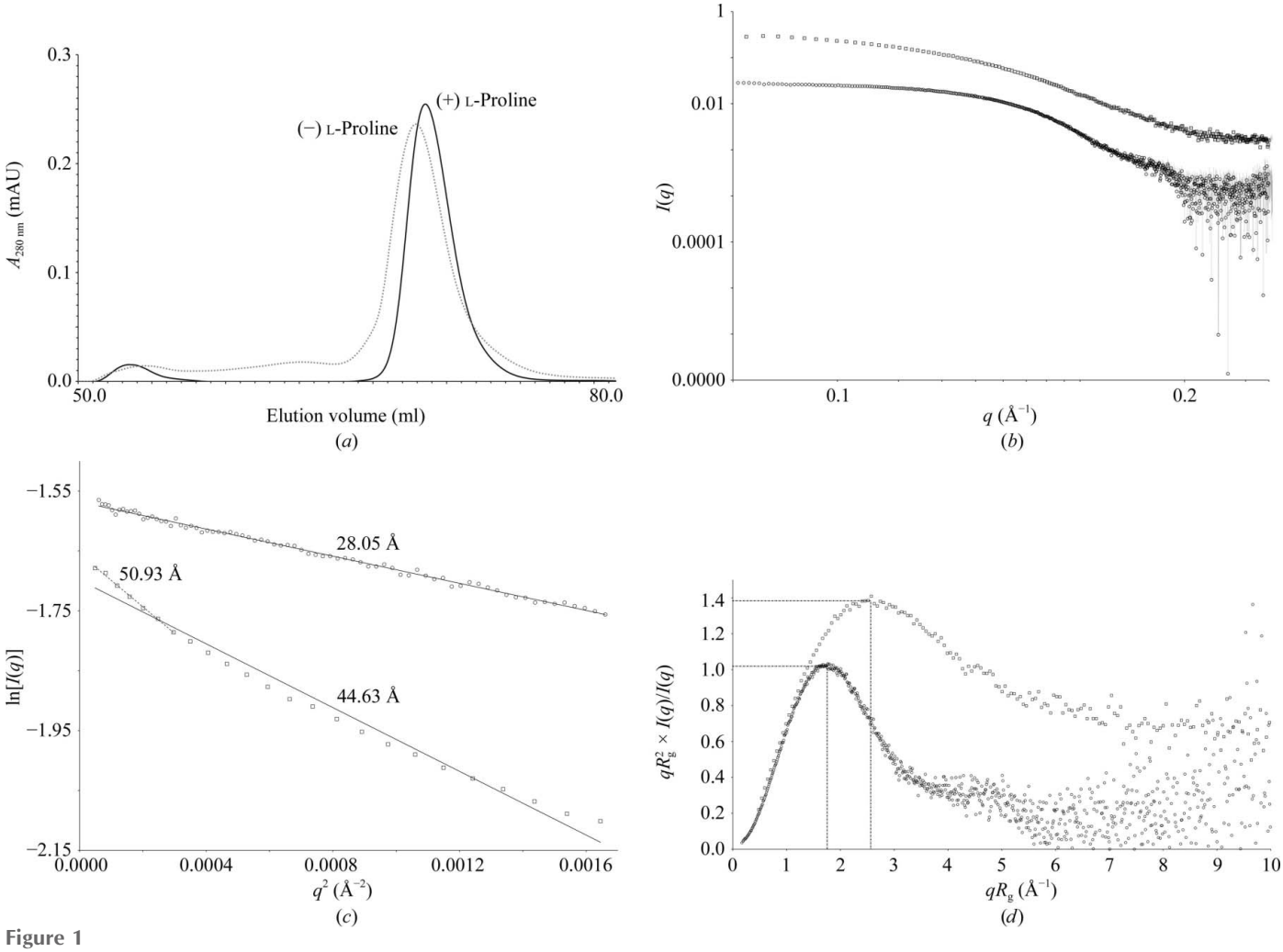


Figure 1
Characterization of the NONO homodimer in solution using small-angle X-ray scattering. (*a*) Size-exclusion chromatography profile for NONO homodimer purified in the presence (solid line) and absence (dashed line) of L-proline. The absorbance at 280 nm is plotted against the elution volume from the S200 16/60 gel-filtration column. (*b*) Comparison of log(*I*) *versus q* small-angle X-ray scattering profiles of *H. sapiens* NONO purified in the presence (empty circles) or absence (empty squares) of L-proline. (*c*) Guinier plot of the low-*q* scattering data indicating the substantially larger radius of gyration (R_g) in the absence of L-proline (44.63 Å, empty squares) compared with the presence of L-proline (28.05 Å, empty circles). An additional fit to the lowest-*q* data is shown for NONO in the absence of L-proline. (*d*) Normalized (dimensionless) Kratky plot of NONO in the presence (empty circles) or absence (empty squares) of L-proline. The peak qR_g^2 for each sample is indicated with a dashed line.

the NONO homodimer, producing a sharper analytical gel-filtration profile (Fig. 1*a*). To further investigate this, we explored small-angle X-ray scattering data collected over the course of optimization, where the NONO homodimer was purified in the presence or the absence of L-proline (Figs. 1*b*, 1*c* and 1*d*). The experimental scattering curves are shown in Fig. 1(*b*), from which Guinier analysis enabled the radius of gyration (R_g) to be determined (Fig. 1*c*). The R_g of the NONO homodimer in the presence of L-proline is 28.05 Å and the normalized (dimensionless) Kratky plot has a peak at 1.02 and a qR_g of 1.73 Å^{-1} (*i.e.* $3^{1/2}$), reflecting a simple globular particle (Durand *et al.*, 2010; Receveur-Brechot & Durand, 2012; Fig. 1*d*). In contrast, NONO purified in the absence of L-proline shows poor linearity in the Guinier region with significant sample aggregation, as indicated by the low-q data giving an R_g of 50.93 Å *versus* 44.63 Å for the same Guinier region identified for NONO in the presence of L-proline (Fig. 1*c*). Furthermore, the normalized Kratky plot has a peak at 1.39 and a qR_g of 2.63 Å^{-1}, reflecting either particle flexibility/disorder or asymmetry (Fig. 1*d*). Collectively, these data suggest that the presence of L-proline assists in maintaining a well ordered and globular fold for the NONO homodimer.

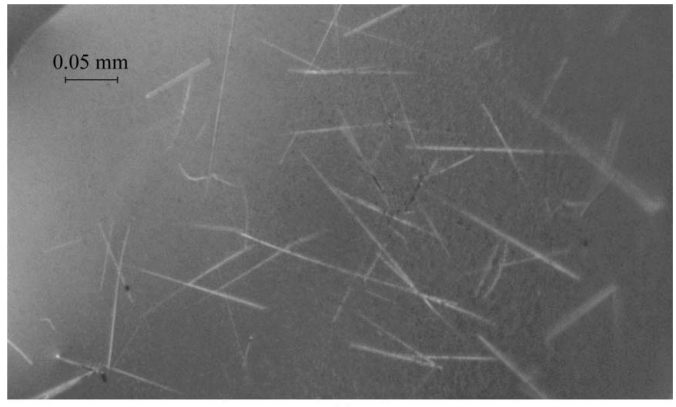
3.2. Crystallization and X-ray data processing

With the addition of L-proline, NONO crystallized over a broad spectrum of PEG-based conditions, with optimization producing large crystalline clusters, in contrast to the thin nondiffracting needles obtained in the absence of L-proline (Figs. 2*b* and 2*a*, respectively). Using the first data-collection strategy, crystals of the NONO homodimer diffracted to beyond 3.0 Å resolution and exhibited a long axis of approximately 400 Å, producing tightly spaced reflections on the detector (Fig. 3*a*) owing to the orientation of the crystal with respect to the beam. The reflections show a distinct 'blurring' coincident with the lines of the crystal lattice, which is particularly evident in the φ = 180° diffraction image (Fig. 3*a*). Data set 1 was initially processed in $P2_1$ (data set 1a; Table 2), where the cumulative intensity distribution L statistic (0.423) indicated twinning (Fig. 3*c*). However, this did not hinder structure solution by molecular replacement. Refinement in $P2_1$ converged to an R_{work} and R_{free} of 21.1 and 24.9%, respectively, with *BUSTER*. To explore the potential twinning, data set 1 was reprocessed in $P1$ (data set 1b; Table 2) and *phenix.xtriage* analysis indicated that the crystal was twinned, with the twin domain operating with a twofold rotation ($-h$, $-k$, l in the $P1$ setting), which together with a pseudo-translation could give rise to $P2_1$. However, the data refined best in $P2_1$ using space-group validation in *ZANUDA*, indicating that the space group was correctly assigned as $P2_1$. Intriguingly, *PHENIX* restrained refinement of data set 1b using the twin operator $-h$, $-k$, l was able to refine the twin fraction to 50% and decrease R_{work} to 19.4% (R_{free} = 23.3%) compared with no twin refinement (R_{work} and R_{free} of 22.3 and 27.9%, respectively).

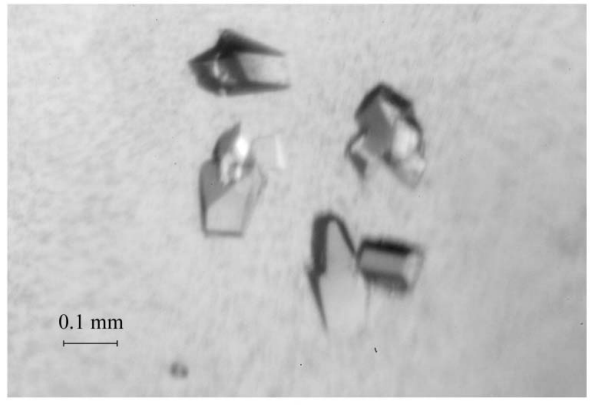
3.3. Data twinning and spatial overlap

While the apparent twinning did not hinder structure solution or refinement, we were interested to find out whether data set 1 had indeed been twinned. After careful analysis of the diffraction images, it became evident that reflections along the long cell axis (~400 Å) were heavily overlapping (Fig. 3*a*; φ = 90°). Thus, it was probable that the presence of indexed overlapping reflections obscured the cumulative intensity distribution. To further evaluate this, we reprocessed data set 1 in *iMosflm* (Battye *et al.*, 2011) with a fixed reduced box size and mosaicity for spot integration and removed the frames suffering from the most severe overlaps (data set 1c; Table 2). While this strategy saw a marked decrease in the data completeness (from 99.8 to 72.4%), the L-test statistic indicated that the data treated in this way were not twinned (L statistic of 0.50; Fig. 3*d*).

To reduce the overlaps and apparent twinning observed in the NONO crystal, a second data set (data set 2) was collected from a different crystal with higher energy X-rays (15 keV; λ = 0.82 Å), a finely collimated beam (10 µm) and an increased crystal-to-detector distance (450 mm, edge at 2.46 Å; Fig. 3*b*). This approach generated well separated reflections on the detector without significantly compromising the resolution range, in spite of the poor crystal orientation with respect to



(*a*)



(*b*)

Figure 2
Crystals of the human NONO homodimer. (*a*) Needle-like crystals of *H. sapiens* NONO grown prior to the addition of L-proline as a buffer component. (*b*) *H. sapiens* NONO crystals containing L-proline with dimensions of approximately 0.1 × 0.05 × 0.1 mm before separation with tools and cryoprotection.

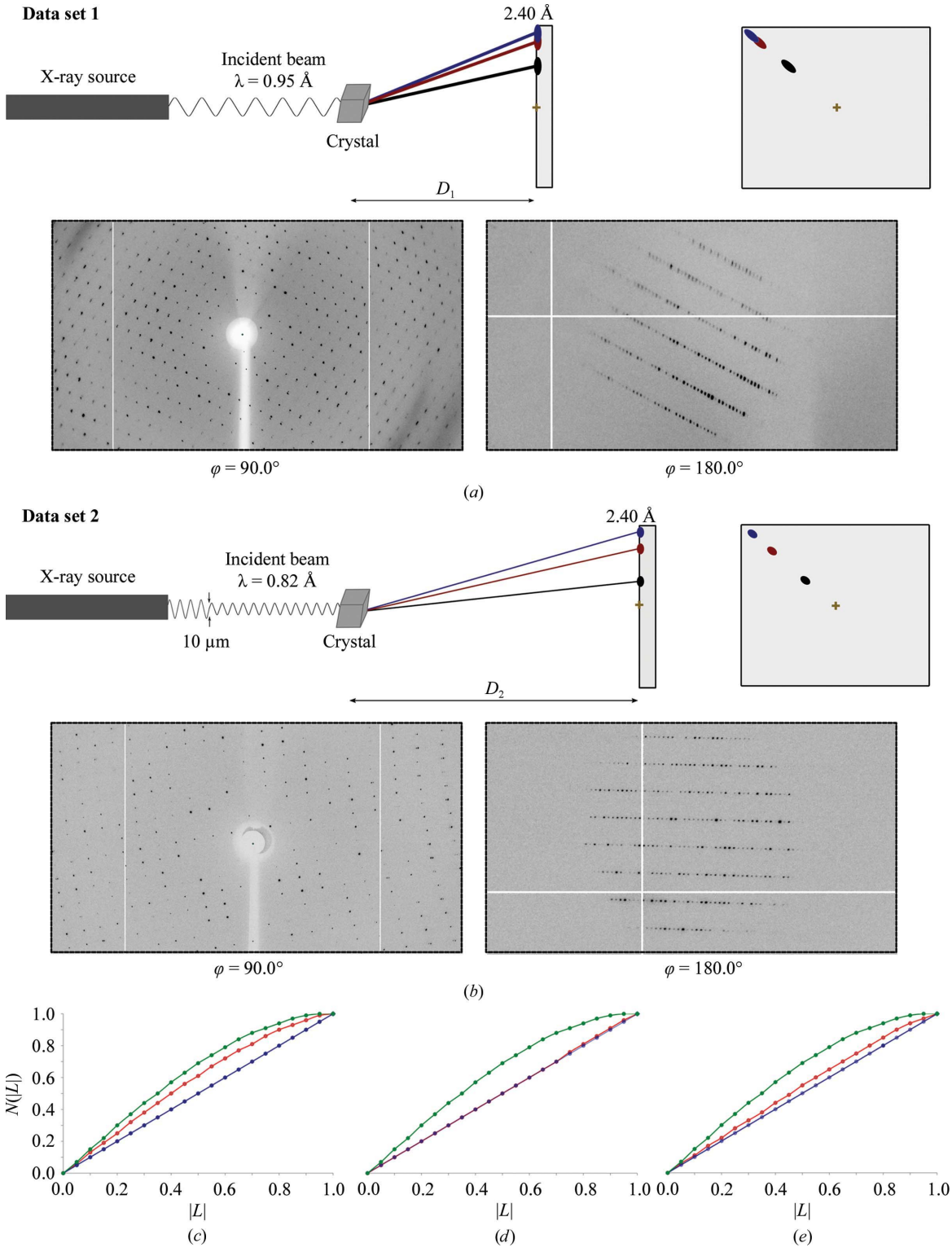


Figure 3
Data-collection strategies employed for the data sets collected in this study with diffraction patterns and the resulting intensity distribution of L. (*a*) Schematic representation of reflections from a crystal collected at 13 keV (λ = 0.95 Å) at a crystal-to-detector distance D_1 (edge at 2.40 Å; data set 1). The oblique overlapping reflections are indicated on the detector. Two representative X-ray diffraction patterns ($\varphi = 90°$ and $\varphi = 180°$) obtained on the MX2 beamline of the Australian Synchrotron at 13 keV using a 1 s exposure over a 0.5° oscillation range are shown below the schematic, where the distinct 'blurring' of spots is evident in the diffraction. (*b*) Schematic representation of reflections collected at 15 keV (λ = 0.82 Å) from a beam collimated to 10 µm over an increased crystal-to-detector distance D_2 (edge at 2.40 Å; data set 2). The less oblique and non-overlapping reflections are indicated on the detector. Two representative X-ray diffraction patterns ($\varphi = 90°$ and $\varphi = 180°$) obtained on the MX2 beamline of the Australian Synchrotron at 15 keV using a 1 s exposure over a 0.5° oscillation range are shown below the schematic. (*c*, *d*, *e*) The intensity distribution of L for theoretically untwinned (blue) and perfectly twinned (green) crystals plotted with the observed intensity distribution for data set 1a (*c*), data set 1c (*d*) and data set 2 (*e*) highlighting the effect of the different data-collection strategies or stringent integration parameters.

the beam. Processing the data in $P2_1$ proceeded with little to no evidence of twinning (Table 2; Fig. 3*e*). The solved structure was refined and the coordinates were deposited in the PDB (R_{work} and R_{free} of 19.7 and 23.4%, respectively; PDB entry 5ifm). However, we note the difficulty of attributing the improvement to any one variable, especially when multiple likely contributing variables were changed (for example, the wavelength, detector distance, collimation and crystal). While the use of higher energy X-rays with an increased crystal-to-detector distance may not significantly affect the size and spatial resolution of spots (Gonzalez *et al.*, 1994), the reduced beam size possibly eliminated the previously described spot 'blurring' within data set 1. Nevertheless, differences in crystal quality between data set 1 and data set 2, possibly owing to post-crystallization handling (*e.g.* cryoprotection), could have similarly improved the general data quality. Further controlled experiments would be necessary to evaluate the variables that actually contribute to the improved data quality.

Evidently, treating the overlapped data (data set 1) with an applied twin operator in *phenix.refine* systematically improved the R_{work} and R_{free} refinement statistics. While this could be attributed to a lower observable:parameter ratio, it is more likely to be a statistical result of comparing intensities (or amplitudes) from a narrower distribution (as is the case when twinning is present; Redinbo & Yeates, 1993; Lebedev *et al.*, 2006). To explore this possibility, we processed both data set 1 and data set 2 (40.0–2.95 Å) in *P*1 and generated a test model from data set 2 in *P*1 with all heteroatoms removed. We then

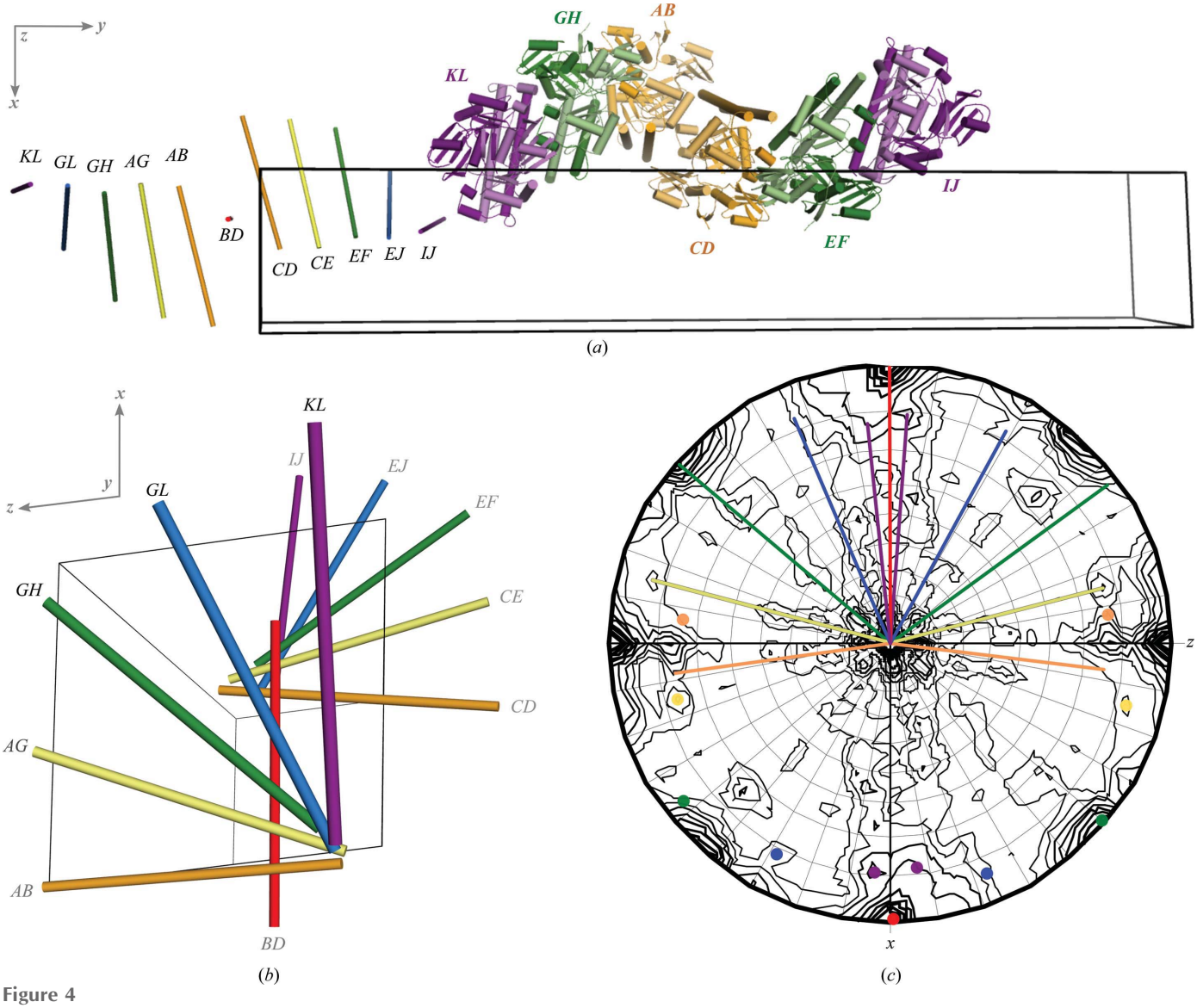


Figure 4
The asymmetric unit content of the X-ray crystal structure of the NONO homodimer (PDB entry 5ifm) illustrated with the extensive noncrystallographic symmetry (NCS). (*a*) NONO dimers are represented as cartoons projected along *z* within the asymmetric unit of the unit cell. The colouring of each NONO dimer reflects NCS within the asymmetric unit. Chain IDs are indicated by single letters. Corresponding paired letters indicate dimers and the vectors describing the observed NCS. (*b*) NCS vectors as in (*a*) viewed down the unique axis. (*c*) $\kappa = 180^\circ$ section of the self-rotation function viewed down the unique axis. Coloured lines represent twofold NCS axes. One line is drawn per NCS vector to emphasize the relation to the molecular packing. Symmetry-related peaks for each vector are represented by dots of the same colour.

rigid-body refined the model against both data sets in *phenix.refine* with and without the twin operator $-h$, $-k$, l. Data set 2 (non-overlapping reflections) gave an R_{work} and R_{free} of 22.12 and 22.27%, respectively, for both twinned and untwinned refinements. However, data set 1 refined to R_{work} and R_{free} values of 22.71 and 22.95%, respectively, when applying the twin refinement and of 23.81 and 23.97%, respectively, when refined without twinning. Collectively, this indicated that the increased parameterization of twin refinement did not artificially decrease the refinement statistics for the non-overlapped data, but caused a significant reduction in R_{work} and R_{free} for the overlapped data. However, the reason for and the broader applicability of this observation remain unclear.

3.4. Structure solution and noncrystallographic symmetry (NCS)

The content of the asymmetric unit of DBHS protein structures solved to date has consistently been a single dimer (Passon *et al.*, 2012; Knott *et al.*, 2015; Lee *et al.*, 2015), with the exception of the pseudosymmetric crystals described for PSPC1–NONO (Lee *et al.*, 2011). For the NONO homodimer, the asymmetric unit consists of six dimers oriented along the z axis in a superhelical array (Fig. 4*a*). The arrangement of NONO dimers can be described as an array of six dimers related by a number of noncrystallographic twofold-rotation axes. Each of the uniquely coloured dimers (*e.g.* the dimers defined as chains *KL* and *GH*) are related within the asymmetric unit by a twofold rotation (blue vector, *GL*). Additionally, the set of three uniquely coloured dimers (*KL*, *GH* and *AB*) are related by a twofold noncrystallographic rotation with a pseudotranslation (red vector, *BD*) to the identically coloured dimers (*CD*, *EF* and *IJ*). The complete set of twofold axes describing the dimer noncrystallographic symmetry can be visualized along the unique axis (y) (Fig. 4*b*). Furthermore, the vectors can be superposed onto the $\kappa = 180^{\circ}$ section of the self-rotation function (Fig. 4*c*), where they coincide with peaks. While the crystal packing observed in the structure of the NONO homodimer is unique, it is unclear whether the superhelical array of NONO dimers has any bearing on the role of DBHS proteins in macromolecular complex formation. The solution scattering of NONO in the presence of L-proline is consistent with the size and shape of a single crystallographically determined NONO homodimer ($\chi = 1.04$), whereas a model consisting of the asymmetric unit content (dodecamer) fits poorly ($\chi > 100$; Fig. 5). Although functional aggregation features prominently in the function of DBHS and paraspeckle protein (Hennig *et al.*, 2015; Lee *et al.*, 2015), it is clear from solution scattering that there is no relationship between the observed NCS and the solution structure. However, we note that the absence of both the coiled-coil oligomerization site (Lee *et al.*, 2015) and a potential nucleic acid scaffold would limit any functional aggregation in this context. Furthermore, the instability of NONO in the absence of L-proline could be suggestive of an underlying structural dynamic that is subject to small-molecule regulation. The preliminary results described here represent a significant first step towards understanding the multifunctionality of NONO and pave the way for future studies of intermolecular interactions.

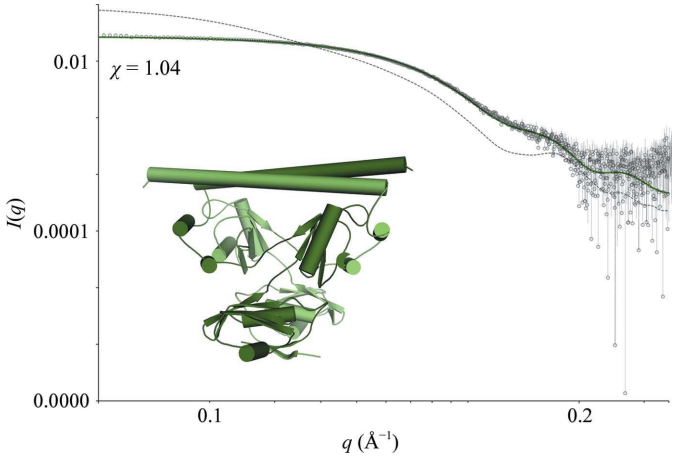


Figure 5
Identification of the solution state of the human NONO homodimer by SAXS. The log(I) *versus* q small-angle X-ray scattering of *H. sapiens* NONO purified in the presence of L-proline (empty circles) is plotted in arbitrary units against q. The theoretical scattering calculated from the X-ray crystal structure of a single NONO homodimer (solid green line; PDB entry 5ifm; chain *AB*) and the asymmetric unit dodecamer (dashed grey line) is fitted to the observed scattering data. A representative cartoon of the NONO homodimer is shown with each chain coloured a different shade of green.

4. Conclusions

The pursuit of the X-ray crystal structure of the NONO homodimer presents a problematic case where there is a need to overcome barriers in both protein purification and X-ray data analysis. Here, by combining solution scattering with crystallography, it was possible not only to identify an additive that improved protein behaviour but to reveal the true solution structure in the light of extensive confounding noncrystallographic symmetry. We also reason that data set 1 was not twinned, but rather the presence of overlapping reflections obscured the L-test statistic. Most typical data-processing software pipelines at synchrotrons flag an abnormal L-test as potential evidence for twinning. In our example it appears instead to be a result of poor spot separation, and so required close inspection of the original images to be resolved, rather than post-processing treatment for twinning. While the processing strategies employed here did enable the deconvolution of crystal twinning and spatial overlap, the importance (in these days of high-throughput data collection) of scrutinizing the diffraction at the beamline cannot be understated, especially in instances of large unit-cell dimensions and crystal-to-crystal variability.

Acknowledgements

This work was supported by National Health and Medical Research Council (NHMRC) Project Grants 1048659 and 1050585 to CSB and AHF, Australian Research Council

Discovery Project DP160102435, the University of Western Australia Hackett Postgraduate Scholarship (GJK) and the European Union FP7 Marie Curie IEF program (AT). Data collection was undertaken on the MX2 and SAXS/WAXS beamlines of the Australian Synchrotron, Victoria, Australia. We are grateful to Andrey Lebedev for helpful discussions, and to Nathan Cowieson, Nigel Kirby and Benjamin Gully for assistance with SAXS data collection.

References

Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst.* D**68**, 352–367.

Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* D**67**, 271–281.

Blanc, E., Roversi, P., Vonrhein, C., Flensburg, C., Lea, S. M. & Bricogne, G. (2004). *Acta Cryst.* D**60**, 2210–2221.

Bond, C. S. (2003). *J. Appl. Cryst.* **36**, 350–351.

Brookes, E., Pérez, J., Cardinali, B., Profumo, A., Vachette, P. & Rocco, M. (2013). *J. Appl. Cryst.* **46**, 1823–1833.

Cao, S., Moss, W., O'Grady, T., Concha, M., Strong, M. J., Wang, X., Yu, Y., Baddoo, M., Zhang, K., Fewell, C., Lin, Z., Dong, Y. & Flemington, E. K. (2015). *J. Virol.* **89**, 7120–7132.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.

Chen, L.-L. & Carmichael, G. G. (2009). *Mol. Cell*, **35**, 467–478.

Dong, B., Horowitz, D. S., Kobayashi, R. & Krainer, A. R. (1993). *Nucleic Acids Res.* **21**, 4085–4092.

Dong, X., Shylnova, O., Challis, J. R. G. & Lye, S. J. (2005). *J. Biol. Chem.* **280**, 13329–13340.

Durand, D., Vivès, C., Cannella, D., Pérez, J., Pebay-Peyroula, E., Vachette, P. & Fieschi, F. (2010). *J. Struct. Biol.* **169**, 45–53.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.

Evans, P. R. & Murshudov, G. N. (2013). *Acta Cryst.* D**69**, 1204–1214.

Fox, A. H., Bond, C. S. & Lamond, A. I. (2005). *Mol. Biol. Cell*, **16**, 5304–5315.

Fox, A. H., Lam, Y. W., Leung, A. K. L., Lyon, C. E., Andersen, J., Mann, M. & Lamond, A. I. (2002). *Curr. Biol.* **12**, 13–25.

Golovanov, A. P., Hautbergue, G. M., Wilson, S. A. & Lian, L.-Y. (2004). *J. Am. Chem. Soc.* **126**, 8933–8939.

Gonzalez, A., Denny, R. & Nave, C. (1994). *Acta Cryst.* D**50**, 276–282.

Gully, B. S., Cowieson, N., Stanley, W. A., Shearston, K., Small, I. D., Barkan, A. & Bond, C. S. (2015). *Nucleic Acids Res.* **43**, 1918–1926.

Hall-Pogar, T., Liang, S., Hague, L. K. & Lutz, C. S. (2007). *RNA*, **13**, 1103–1115.

Hennig, S. *et al.* (2015). *J. Cell Biol.* **210**, 529–539.

Ishitani, K., Yoshida, T., Kitagawa, H., Ohta, H., Nozawa, S. & Kato, S. (2003). *Biochem. Biophys. Res. Commun.* **306**, 660–665.

Izumi, H., McCloskey, A., Shinmyozu, K. & Ohno, M. (2014). *Nucleic Acids Res.* **42**, 3998–4007.

Jancarik, J., Pufan, R., Hong, C., Kim, S.-H. & Kim, R. (2004). *Acta Cryst.* D**60**, 1670–1673.

Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.

Kameoka, S., Duque, P. & Konarska, M. M. (2004). *EMBO J.* **23**, 1782–1791.

Kaneko, S., Rozenblatt-Rosen, O., Meyerson, M. & Manley, J. L. (2007). *Genes Dev.* **21**, 1779–1789.

Kirby, N. M., Mudie, S. T., Hawley, A. M., Cookson, D. J., Mertens, H. D. T., Cowieson, N. & Samardzic-Boban, V. (2013). *J. Appl. Cryst.* **46**, 1670–1680.

Kleywegt, G. J. (1996). *Acta Cryst.* D**52**, 842–857.

Knott, G. J., Bond, C. S. & Fox, A. H. (2016). *Nucleic Acids Res.*, doi:10.1093/nar/gkw271.

Knott, G. J., Lee, M., Passon, D. M., Fox, A. H. & Bond, C. S. (2015). *Protein Sci.* **24**, 2033–2043.

Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. (2003). *J. Appl. Cryst.* **36**, 1277–1282.

Kowalska, E., Ripperger, J. A., Hoegger, D. C., Bruegger, P., Buch, T., Birchler, T., Mueller, A., Albrecht, U., Contaldo, C. & Brown, S. A. (2013). *Proc. Natl Acad. Sci. USA*, **110**, 1592–1599.

Kowalska, E., Ripperger, J. A., Muheim, C., Maier, B., Kurihara, Y., Fox, A. H., Kramer, A. & Brown, S. A. (2012). *Mol. Cell. Biol.* **32**, 4585–4594.

Lebedev, A. A. & Isupov, M. N. (2014). *Acta Cryst.* D**70**, 2430–2443.

Lebedev, A. A., Vagin, A. A. & Murshudov, G. N. (2006). *Acta Cryst.* D**62**, 83–95.

Lee, M., Passon, D. M., Hennig, S., Fox, A. H. & Bond, C. S. (2011). *Acta Cryst.* D**67**, 981–987.

Lee, M., Sadowska, A., Bekere, I., Ho, D., Gully, B. S., Lu, Y., Iyer, K. S., Trewhella, J., Fox, A. H. & Bond, C. S. (2015). *Nucleic Acids Res.* **43**, 3826–3840.

Liang, S. & Lutz, C. S. (2006). *RNA*, **12**, 111–121.

Lu, J. Y. & Sewer, M. B. (2015). *Mol. Cell. Biol.* **35**, 1223–1237.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Mircsof, D. *et al.* (2015). *Nature Neurosci.* **18**, 1731–1736.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.

Padilla, J. E. & Yeates, T. O. (2003). *Acta Cryst.* D**59**, 1124–1130.

Passon, D. M., Lee, M., Fox, A. H. & Bond, C. S. (2011). *Acta Cryst.* F**67**, 1231–1234.

Passon, D. M., Lee, M., Rackham, O., Stanley, W. A., Sadowska, A., Filipovska, A., Fox, A. H. & Bond, C. S. (2012). *Proc. Natl Acad. Sci. USA*, **109**, 4846–4850.

Peng, R., Dye, B. T., Pérez, I., Barnard, D. C., Thompson, A. B. & Patton, J. G. (2002). *RNA*, **8**, 1334–1347.

Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., Gorba, C., Mertens, H. D. T., Konarev, P. V. & Svergun, D. I. (2012). *J. Appl. Cryst.* **45**, 342–350.

Receveur-Brechot, V. & Durand, D. (2012). *Curr. Protein Pept. Sci.* **13**, 55–75.

Redinbo, M. R. & Yeates, T. O. (1993). *Acta Cryst.* D**49**, 375–380.

Seabrook, S. A. & Newman, J. (2013). *ACS Comb. Sci.* **15**, 387–392.

St Gelais, C., Roger, J. & Wu, L. (2015). *AIDS Res. Hum. Retroviruses*, **31**, 806–816.

Semenyuk, A. V. & Svergun, D. I. (1991). *J. Appl. Cryst.* **24**, 537–540.

Straub, T., Grue, P., Uhse, A., Lisby, M., Knudsen, B. R., Tange, T. O., Westergaard, O. & Boege, F. (1998). *J. Biol. Chem.* **273**, 26261–26264.

Svergun, D., Barberato, C. & Koch, M. H. J. (1995). *J. Appl. Cryst.* **28**, 768–773.

Udayakumar, D. & Dynan, W. S. (2015). *Biochem. Biophys. Res. Commun.* **463**, 473–478.

Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.

Yadav, S. P., Hao, H., Yang, H.-J., Kautzmann, M. A., Brooks, M., Nellissery, J., Klocke, B., Seifert, M. & Swaroop, A. (2014). *Hum. Mol. Genet.* **23**, 2132–2144.

Zhu, Z., Zhao, X., Zhao, L., Yang, H., Liu, L., Li, J., Wu, J., Yang, F., Huang, G. & Liu, J. (2016). *Oncogene*, **35**, 1399–1410.